U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
PB-298 056

The Analysis of Suspended Particulates and Sulfates A Way to Begin

Tennessee Valley Authority, Chattanooga

Prepared for

Environmental Protection Agency, Washington, DC Office of Energy, Minerals and Industry

Mar 79

TVA

EPA

Tennessee Valley Authority
Office of Natural Resources
Chattanooga TN 37401

TVA/ONR-79/03

PB 298056

United States Environmental Protection Agency
Office of Energy, Minerals, and Industry
Washington DC 20460

EPA-600/7-79-084
March 1979

Research and Development

The Analysis of Suspended Particulates and Sulfates

A Way to Begin

Interagency Energy/Environment R&D Program Report

REPRODUCED BY
NATIONAL TECHNICAL
INFORMATION SERVICE
U. S. DEPARTMENT OF COMMERCE
SPRINGFIELD, VA. 22161

RESEARCH REPORTING SERIES

Research reports of the Office of Research and Development, U.S. Environmental Protection Agency, have been grouped into nine series. These nine broad categories were established to facilitate further development and application of environmental technology. Elimination of traditional grouping was consciously planned to foster technology transfer and a maximum interface in related fields. The nine series are:

1. Environmental Health Effects Research
2. Environmental Protection Technology
3. Ecological Research
4. Environmental Monitoring
5. Socioeconomic Environmental Studies
6. Scientific and Technical Assessment Reports (STAR)
7. Interagency Energy-Environment Research and Development
8. "Special" Reports
9. Miscellaneous Reports

This report has been assigned to the INTERAGENCY ENERGY-ENVIRONMENT RESEARCH AND DEVELOPMENT series. Reports in this series result from the effort funded under the 17-agency Federal Energy/Environment Research and Development Program. These studies relate to EPA's mission to protect the public health and welfare from adverse effects of pollutants associated with energy systems. The goal of the Program is to assure the rapid development of domestic energy supplies in an environmentally-compatible manner by providing the necessary environmental data and control technology. Investigations include analyses of the transport of energy-related pollutants and their health and ecological effects; assessments of, and development of, control technologies for energy systems; and integrated assessments of a wide range of energy-related environmental issues.

This document is available to the public through the National Technical Information Service, Springfield, Virginia 22161.

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/7-79-084	2.
3. RECIPIENT'S ACCESSION NO. PB298056	
4. TITLE AND SUBTITLE THE ANALYSIS OF SUSPENDED PARTICULATES AND SULFATES: A WAY TO BEGIN	5. REPORT DATE March 1979
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Walter Liggett and William Parkhurst	8. PERFORMING ORGANIZATION REPORT NO. TVA/ONR-79/03
9. PERFORMING ORGANIZATION NAME AND ADDRESS Office of Natural Resources Tennessee Valley Authority Chattanooga, TN 37401	10. PROGRAM ELEMENT NO. INE - 625 B 11. CONTRACT/GRANT NO. 80 BDM
12. SPONSORING AGENCY NAME AND ADDRESS U.S. Environmental Protection Agency Office of Research & Development Office of Energy, Minerals & Industry Washington, D.C. 20460	13. TYPE OF REPORT AND PERIOD COVERED Milestone 14. SPONSORING AGENCY CODE EPA/600/17

15. SUPPLEMENTARY NOTES
This project is part of the EPA-planned and coordinated Federal Interagency Energy/Environment R&D Program.

16. ABSTRACT

Total suspended particulate (TSP) and suspended sulfate (SS) levels have been sampled since November 1973 at five isolated sites across the Tennessee Valley. A method for beginning to analyze such data is demonstrated. This beginning is intended to lead finally to information on pollution sources, an objective that may require modeling meteorological influences and resolving sources. Analysis with this objective, which can be very complex, is effectively begun by using the method demonstrated in this paper. Applied to the TSP and SS data, this method suggests agricultural contributions to TSP levels, distant-source contributions to SS levels, and various influences of the meteorology. This method also shows deficiencies in the data collection that prevent the building of better, more quantitative models. One deficiency in this data set is the sixth-day sampling, which is not frequent enough to allow monthly variations in pollution levels to be distinguished from more rapid variations. Thus, data analysis would be more effective if the sampling frequency were increased and, further, if particle size and chemical composition were better resolved.

17. (Circle One or More) KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group
Inorganic Chemistry	Charac. Meas. & Monit.	7B

18. DISTRIBUTION STATEMENT Release to public	19. SECURITY CLASS (This Report) Unclassified	21. NO. OF PAGES 32
	20. SECURITY CLASS (This page) Unclassified	22. PRICE PC A03/MF A01

EPA Form 2220-1 (9-73)

EPA-600/7-79-084
TVA/ONR-79/03
March 1979

THE ANALYSIS OF SUSPENDED PARTICULATES AND SULFATES: A WAY TO BEGIN

by

Walter Liggett and William Parkhurst
Office of Natural Resources
Tennessee Valley Authority
Chattanooga, Tennessee 37401

Interagency Agreement No. EPA-IAG-D5-E721
Project No. 80 BDM
Program Element No. INE-625B

Project Officer

James T. Stemmle
401 M Street RD-681
Washington, D.C. 20460

Prepared for

OFFICE OF ENERGY, MINERALS, AND INDUSTRY
OFFICE OF RESEARCH AND DEVELOPMENT
U.S. ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

DISCLAIMER

This report was prepared by the Tennessee Valley Authority and has been reviewed by the Office of Energy, Minerals, and Industry, U.S. Environmental Protection Agency, and approved for publication. Approval does not signify that the contents necessarily reflect the views and policies of the Tennessee Valley Authority or the U.S. Environmental Protection Agency, nor does mention of trade names or commercial products constitute endorsement or recommendation for use.

ABSTRACT

Total suspended particulate (TSP) and suspended sulfate (SS) levels have been sampled since November 1973 at five isolated sites across the Tennessee Valley. A method for beginning to analyze such data is demonstrated. This beginning is intended to lead finally to information on pollution sources, an objective that may require modeling meteorological influences and resolving sources. Analysis with this objective, which can be very complex, is effectively begun by using the method demonstrated in this paper. Applied to the TSP and SS data, this method suggests agricultural contributions to TSP levels, distant-source contributions to SS levels, and various influences of the meteorology. This method also shows deficiencies in the data collection that prevent the building of better, more quantitative models. One deficiency in this data set is the sixth-day sampling, which is not frequent enough to allow monthly variations in pollution levels to be distinguished from more rapid variations. Thus, data analysis would be more effective if the sampling frequency were increased and, further, if particle size and chemical composition were better resolved.

This report was submitted by the Tennessee Valley Authority, Office of Natural Resources, in partial fulfillment of Energy Accomplishment Plan 80 BDM under terms of Interagency Agreement EPA-IAG-D5-E721 with the Environmental Protection Agency. Work was completed as of January 12, 1979.

CONTENTS

	Page
Abstract	iii
List of Figures	v
List of Tables	v
1. Introduction	1
2. Conclusions and Recommendations	3
3. The Method	4
Overview of the method	4
The TSP and SS components	5
The algorithm	7
4. Interpretation of the Data	14
Seasonal component	14
Valley-wide and local smooths	15
Valley-wide and individual roughs	17
5. Design of Monitoring	20
References	22

Preceding page blank

LIST OF FIGURES

Number		Page
1	Seasonal patterns for TSP and SS	6
2	Smoothed levels of total suspended particulates . .	8
3	Smoothed levels of suspended sulfates	9
4	Data decomposition showing flow of calculations . .	11
5	Daily sulfate data with sixth-day sampling smoothed	16

LIST OF TABLES

Number		Page
1	Robust Correlations and Number of Nonmissing Observations for the Roughs	10

SECTION 1

INTRODUCTION

Since November 1973, the Tennessee Valley Authority (TVA) has operated high-volume samplers at five sites to obtain background concentrations and trends for total suspended particulates (TSP) and water-soluble suspended sulfates (SS). These sites, which are intended to represent large subregions of the Tennessee Valley, are remote from power plants and other large sources of industrial pollution. From east to west, these sites are in Washington County, Virginia (at Loves Mill); Monroe County, Tennessee (at Loudon); Jackson County, Alabama (at Hytop); Giles County, Tennessee; and Trigg County, Kentucky [at Land Between The Lakes (LBL)]. Samples have been collected for a 24-h period every sixth day and analyzed by standard methods.[1-4]

This paper demonstrates a method that helps investigators explain data like these. The explanations answer questions such as how much each source contributed to the observed levels, an important question in the application of the 1977 Clean Air Act Amendments.[5] The method may suggest explanations with clear implications. However, the method may be only the first step in developing a more complete model of what influences the measurements. In this case, the method is intended to show the potential benefits of a more complete model and the requirements for its development so that the considerable expense and expertise possibly needed can be justified and planned. Important benefits may not be available from a particular data set because meteorological influences or something else cannot be adequately modeled. The method is intended to indicate such a possibility.

Models that explain air quality measurements are needed for many purposes, for example, to obtain information relevant to control strategies or to interpret the trends that monitoring is meant to detect.[6] Such models involve several factors including the sources of pollution and the transport and transformation of pollutants. Such models are needed because they differentiate among these factors. Thus, they allow the effects of control strategies and other changes to be predicted and the causes of a trend to be understood.

The method demonstrated in this paper decomposes the data into components that represent data variations of different temporal and spatial extents.[7,8] A guide to the method is given by the equation,

log-transformed data = seasonal component +
Valley-wide smooth + local smooth + individual rough.

This equation shows that log transforms of the data rather than the original data are decomposed into components. One component represents the seasonal (i.e., annually recurring) variation for all sites. Smooths, which show variations that persist in time, are computed for all sites (Valley-wide) and for the variations unique to each site (local). The individual roughs are the irregular variations not accounted for by the other components. A Valley-wide rough has been computed, but not incorporated in the decomposition.

The method is useful because the data are easier to interpret component by component than all at once. The data are determined by many factors. The influence of these factors on each component is easier to understand than their influence on the undivided data. For example, the influence of seasonal factors can be seen in the seasonal component, but generally not in the other components. Thus, the method is much more revealing, yet no more complicated, than the histograms often used to summarize air quality data.

The TSP and SS data from remote sites on which the method is demonstrated are interesting because of questions about pollutant origins. These origins are both distant sources and local, nonindustrial sources such as agriculture. The questions involve the methods and benefits of controlling such sources and the interference of such sources with the monitoring of a specific industrial source.

Decomposing these data into components reveals several features observed in other regions. One feature is the patterns shown by the seasonal component. Nationwide, seasonal patterns in TSP are not consistent, indicating the importance of local sources, which differ for urban and rural monitoring.[9] In the east, the seasonal patterns in SS have a single peak in the summer.[9] Another feature is the relations among the series observed at different sites. For SS data, similarities in time behavior at widely separated sites have been observed in New York State.[10-12] Some intersite differences observed in the data analyzed here (unusually high SS levels at LBL) have been explained by Reisinger and Crawford.[13]

To describe the method, we discuss the components it produces from the TSP and SS data before we specify the computational details. This discussion, which is in Section 3, is thus more data-oriented than the usual description of a method. In Sections 4 and 5, we interpret the data presented in Section 3. Section 4 discusses the physical mechanisms responsible for the observations. Section 5 discusses consequences for the design of monitoring.

SECTION 2

CONCLUSIONS AND RECOMMENDATIONS

Many factors other than emission levels influence air quality monitoring data; most obviously, the weather influences transport. Further, many sources other than those usually controlled contribute to pollution levels. We recommend that, when possible, the influences of these factors be modeled rather than treated as random.

The influences on pollution levels most obvious in monitoring data are often of little interest in decision making. When this is true, we recommend that the data collection and analysis needed to adjust for these influences be undertaken. For example, adjustment of the data for meteorological influences should allow emission trends to be detected more easily.

The effort finally needed to model the influences on the data might require expertise and data collection, which make monitoring much more expensive. We recommend that the importance of the information to be obtained determine the degree of monitoring to be done.

The analysis method demonstrated helps guide future data collection and analysis by showing what is needed to meet objectives. We recommend that all monitoring data be subjected to such preliminary analysis.

SECTION 3

THE METHOD

OVERVIEW OF THE METHOD

Both the TSP and the SS data are composed of time series from each site. These series each have 258 observations that have been transformed by:

$$y = \log_{10}(x + 1), \tag{1}$$

where

x = an original observation, $\mu g/m^3$,

y = the corresponding log-transformed observation, $\mu g/m^3$.

Data transformations are discussed by Tukey.[7]

The first part of the decomposition computes a smooth trace through each series. This trace follows the slowly changing variations in the data, the variations that persist from sample to sample. It is not affected by the irregular sample-to-sample changes. It represents the data variations that monthly averages are intended to portray. Subtracting the smooth trace from the series that generated it gives a component that represents irregular sample-to-sample changes in the data. Thus, each series is decomposed into two components, a smooth trace and an irregular component called an individual rough. The smooth trace represents data fluctuations caused, for example, by seasonal changes in the weather, and the rough represents fluctuations caused, for example, by frontal passages.

The smooth traces are computed by the use of running medians. Consider, for example, a running median that spans five observations. It is computed by finding the middle value (the third largest value) of every group of five successive values of a series. An alternative, a running month-long average, is computed by finding the average of every group of five successive values. A running median is less sensitive to isolated values that are very large or very small. Thus, running medians give a smooth trace that is less influenced by such values and, consequently, a rough that better represents such values. The actual algorithm for computing the smooth traces, which is described below, involves repeated computation of running medians, a method for obtaining the smooth trace at the ends of the series, and an approach to missing values.

The second part of the decomposition extracts the Valley-wide component from the smooth traces for each site. Our choice for the Valley-wide component is the sample-by-sample average of the five smooth traces. This choice was made despite one- and two-day differences in sampling day that occur before May 1976 because smooth traces rather than the original data are averaged. Subtracting the Valley-wide component from the smooth traces gives the local smooths.

The third part of the decomposition extracts the seasonal component from the Valley-wide component. The seasonal component is computed without the first and last seven values of the Valley-wide component so that exactly four years of data are used. Since each year has sixty-one values, the seasonal component has sixty-one values. Each of these values is the midmean of the corresponding four yearly values. (The midmean of four values is the average of the second and third largest.) Subtracting the seasonal component from the Valley-wide component gives the Valley-wide smooth. The Valley-wide smooth shows unusual years more clearly because the midmean instead of the average is used to compute the seasonal component.

THE TSP AND SS COMPONENTS

Further understanding of the method can be gained by considering the components produced from the TSP and SS data. However, before presenting these components, we present annual and 24-h summaries of these data to help the reader relate them to other data.

For the calendar years 1974 through 1977, the annual geometric means of the TSP for these sites ranged from 28 to 43 $\mu g/m^3$. These TSP levels are well below the primary and secondary National Ambient Air Quality Standards of 75 and 60 $\mu g/m^3$, respectively. The 24-h TSP levels found in these data also do not exceed the primary and secondary standards of 150 and 260 $\mu g/m^3$, respectively. However, the 24-h levels for February 24, 1977, a day during a severe dust storm, are recorded as lost records. These levels are actually 88, 767, 699, 654, and 138 $\mu g/m^3$ for Loves Mill, Loudon, Hytop, Giles County, and LBL, respectively, as shown by TVA laboratory files. This dust storm caused 24-h levels to exceed standards throughout the Southeast.[14]

For the same periods and sites, the annual arithmetic means of the SS ranged from 5.9 to 10.0 $\mu g/m^3$. These levels are within the range expected in rural areas east of the Mississippi River.[15] Some states have standards for SS, and the EPA is considering national standards. Suggestions for the annual standard[16] lie between 5 and 15 $\mu g/m^3$, and suggestions for the 24-h standard[16] lie between 10 and 25 $\mu g/m^3$. Fourteen instances of 24-h levels above 25 $\mu g/m^3$ are contained in the data from these sites.

Consider the components discussed above, starting with the seasonal component. The TSP and SS seasonal components are the most pronounced feature of the data. They are shown in Figure 1 after retransformation to compensate for the log transform. They are plotted on a horizontal axis that starts on the first day of winter and is divided seasonally. The estimate of the TSP pattern has peaks in mid-April and mid-July that reach 54 $\mu g/m^3$. It has levels as low as 22 $\mu g/m^3$. The estimate of the SS pattern has a peak in mid-July that reaches 13.0 $\mu g/m^3$. It has levels as low as 3.7 $\mu g/m^3$. The April and July TSP peaks invite comparison because the SS is a much larger fraction of the July TSP peak than of the April TSP peak.

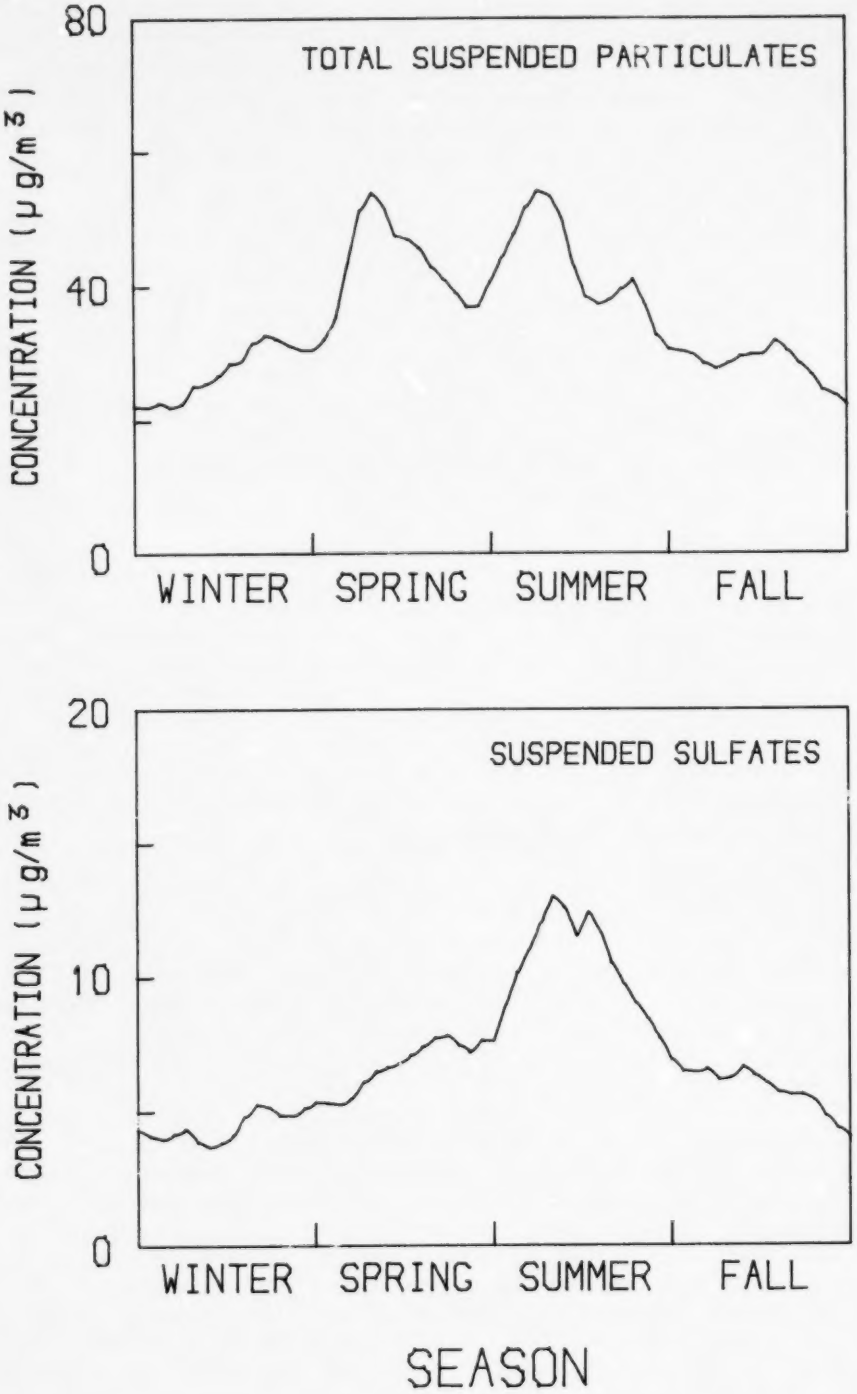

Figure 1. Seasonal patterns for TSP and SS.

The Valley-wide and local smooths in Figure 2 show any annual trends and persistent local conditions contained in the TSP data. The Valley-wide smooth shows that 1974 and 1977 are worse than 1975 and 1976, but it does not seem to provide convincing evidence of an increasing trend. Further, the Valley-wide smooth shows peaks in fall 1974 and in 1977 that invite explanation. The local smooths show that Hytop and Loudon have generally higher levels than the other sites. They also show some interesting peaks.

The corresponding smooths for SS are shown in Figure 3. The Valley-wide smooth seems to show a decreasing trend. As part of this trend, the Valley-wide smooth shows that the winter, spring, and summer of 1975 had unusually high levels. The local smooths show that Hytop has generally higher levels than the other sites and that, except for 1974, Loudon has higher winter levels. Like the TSP smooths, these smooths have many peaks that suggest further investigation.

The roughs are better summarized by the correlations shown in Table 1 than depicted by graphs because the roughs appear nearly random. This table requires four explanations. First, before May 17, 1976, Loves Mill was sampled one day and Loudon was sampled two days before the other three sites. Starting May 17, 1976, all sites were sampled on the same day. Thus, the table has two entries for Loves Mill, Loudon, and the Valley-wide rough, the first for the earlier period and the second for the later period. Second, the Valley-wide rough summarizes the three, then five, roughs from the sites sampled on the same day. The table contains correlations of the individual roughs and the Valley-wide rough to show the similarity of these roughs. Third, the table contains in the lower triangle the numbers of observations not missing and therefore included in the correlations. These numbers are helpful in making inferences. Fourth, the correlations are computed by a robust method that prevents a few observations from dominating the results. This method is the standardized sum and difference method with 5 percent Winsorized variances centered at 10 percent trimmed means.[17]

The roughs have two striking features: (1) Roughs from sites sampled the same day are closely related; and (2) in most cases, for the days on which the Valley-wide rough is unusually high or low, all sites have unusually high or low levels.

THE ALGORITHM

Having described the data, we now show how the decomposition is computed. Before the decomposition is started, missing values in the data are replaced by linear interpolation between nearby values from the same site. Each data point is then transformed as described by Equation (1). These steps produce a 5 x 258 array of values that should be thought of as being in block 1 of Figure 4 at the start of the decomposition. These values represent the period from November 1973 through January 1978. From each smooth, seven values are dropped from each end to reduce the smooths to exactly four years.

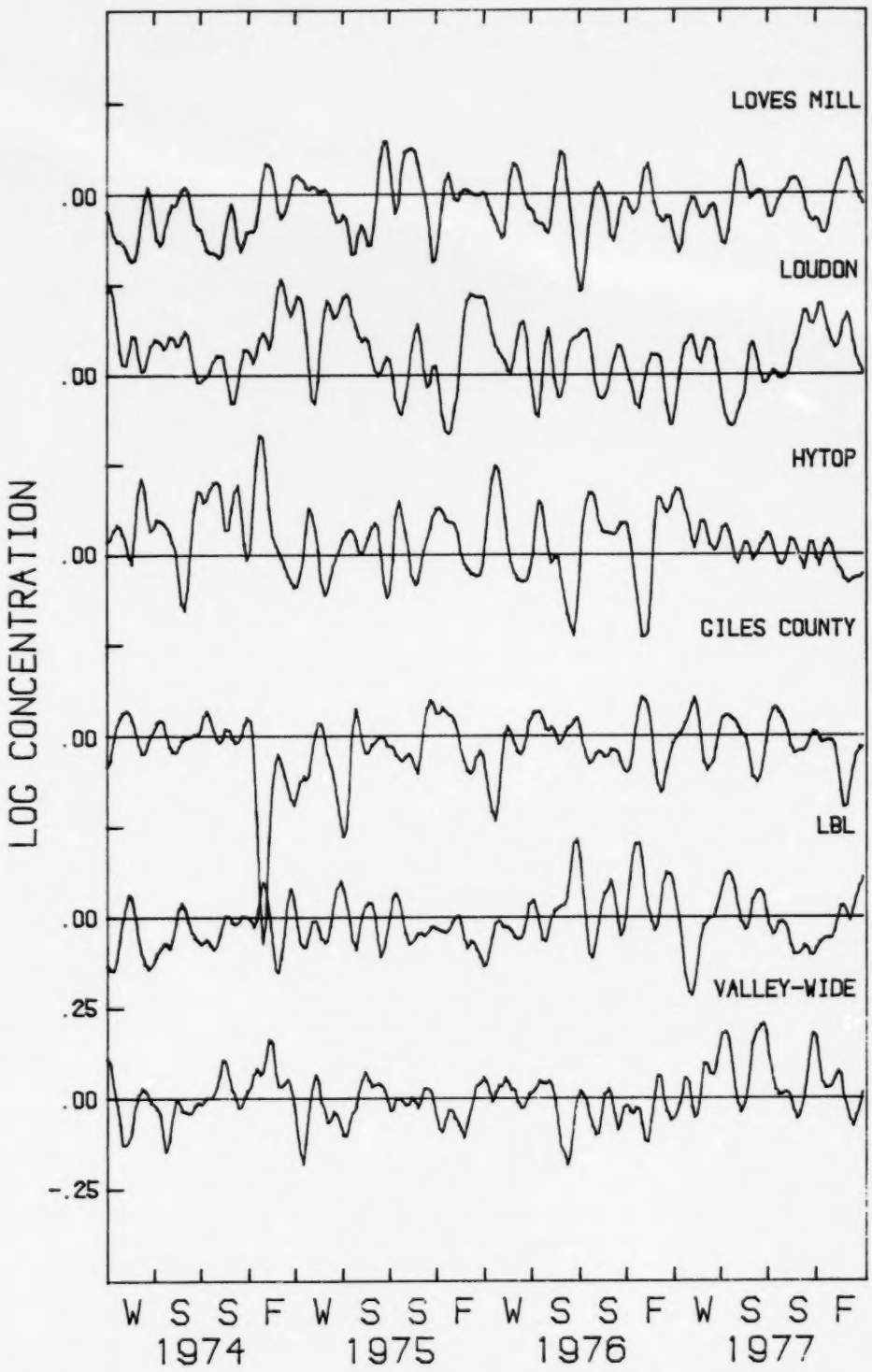

Figure 2. Smoothed levels of total suspended particulates.

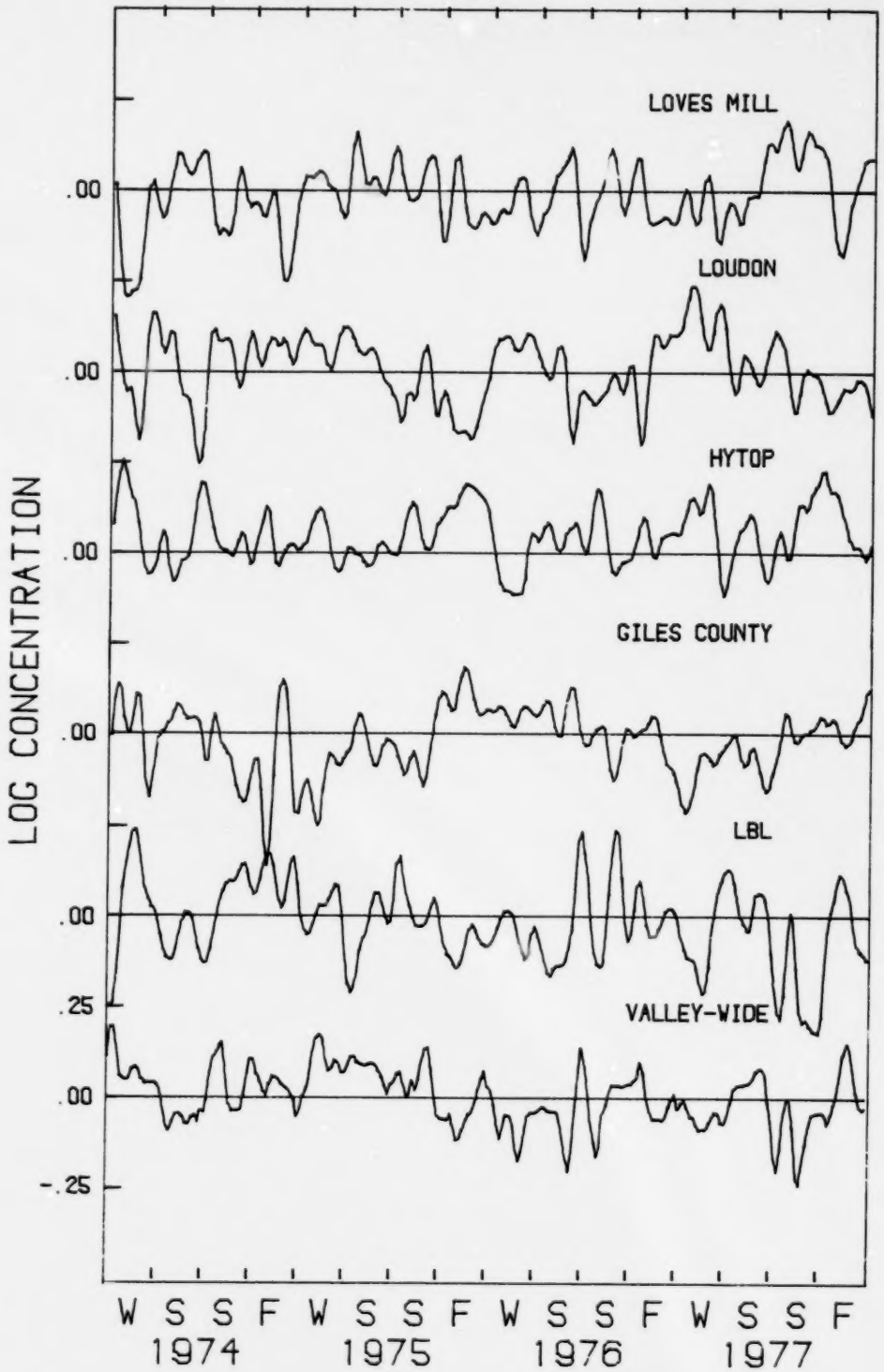

Figure 3. Smoothed levels of suspended sulfates.

TABLE 1. ROBUST CORRELATIONS AND NUMBER OF NONMISSING OBSERVATIONS FOR THE ROUGHS

	Loves Mill	Loudon	Hytop	Giles County	LBL	Valley-wide
			Total suspended particulates			
Loves Mill	--	0.09/0.58[a]	0.32/0.48	0.31/0.57	0.35/0.52	0.34/0.76
Loudon	129/ 84[b]	--	0.12/0.61	-0.01/0.60	0.09/0.51	0.05/0.76
Hytop	132/ 90	128/ 88	--	0.62	0.53	0.84/0.77
Giles County	131/ 79	126/ 77	212	--	0.60	0.83/0.91
LBL	137/ 91	133/ 87	228	217	--	0.84/0.76
Valley-wide	143/ 95	138/ 92	142/ 97	141/ 85	148/ 98	--
			Suspended sulfates			
Loves Mill	--	0.30/0.58[a]	0.43/0.55	0.21/0.54	0.33/0.47	0.40/0.73
Loudon	125/ 86[b]	--	0.09/0.69	0.07/0.58	0.21/0.45	0.11/0.83
Hytop	130/ 92	126/ 89	--	0.66	0.51	0.90/0.83
Giles County	132/ 82	126/ 80	217	--	0.46	0.82/0.83
LBL	133/ 93	129/ 89	227	222	--	0.73/0.67
Valley-wide	141/ 97	136/ 93	142/ 98	144/ 88	145/100	--

[a]Correlation before May 17, 1976/correlation after May 17, 1976.

[b]Number before May 17, 1976/number after May 17, 1976.

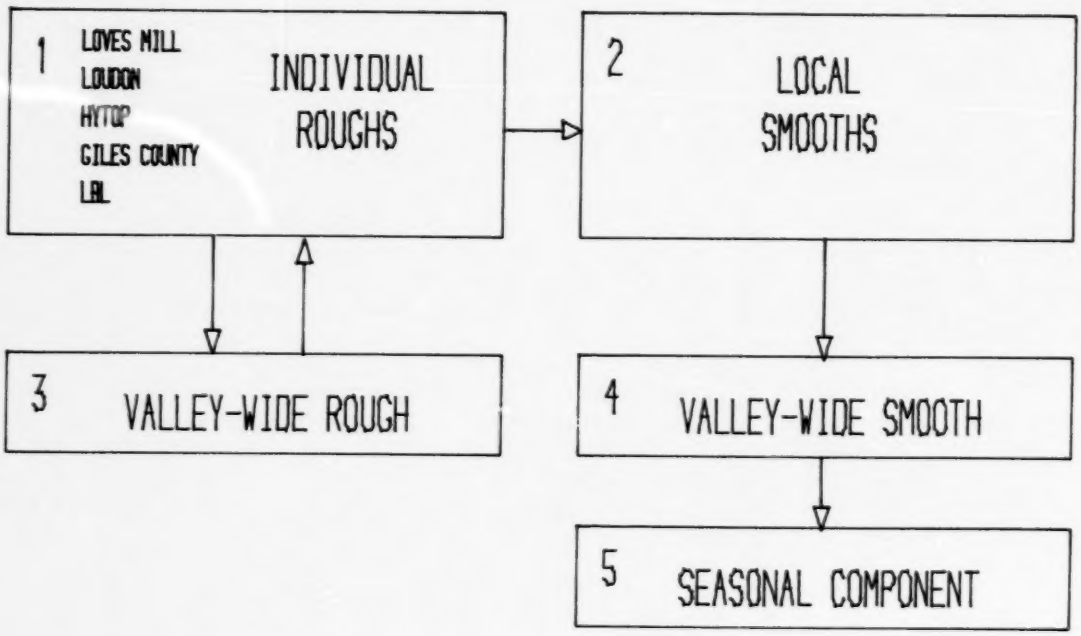

Figure 4. Data decomposition showing flow of calculations.

The first step in the decomposition computes a smooth trace through the data for each site. The particular algorithm we chose for this purpose is called 4253H and is specified below.[8] The smooth traces are subtracted from the data originally in block 1 and stored in block 2; the differences are left in block 1.

The second step computes the Valley-wide component from the values remaining in block 1 by finding the median value for each sampling day. Before May 17, 1976, these medians are determined by Hytop, Giles County, and Land Between the Lakes only. Thereafter, they are determined by all sites. The resulting medians are stored in block 3.

The third step replaces the values stored in block 1 that were initially missing. These values are replaced by the corresponding value of the Valley-wide component in block 3, except for the missing values from Loves Mill and Loudon before May 17, 1976. Missing values from these two sites before May 17, 1976, are replaced by zero.

The fourth step ensures that in the end the values in block 1 have no smooth trace. It repeats computations like those in steps 1 through 3, using the values left in block 1 as inputs. The smooth traces of the values in block 1 are computed, subtracted from the values in block 1, and added to the values in block 2. Next, the Valley-wide component is recomputed as in step 2 and stored in block 3. Then, missing values are replaced as in step 3. Finally, these analogs of steps 1 through 3 are repeated yet another two times. What then remains in block 1 are the individual roughs, and what then remains in block 3 is the Valley-wide rough.

The fifth step removes the Valley-wide component from the values stored in block 2. It averages the values for each sampling day, ignoring the one- and two-day differences in the schedule. These averages are subtracted from block 2, leaving the local smooths, and are stored in block 4.

The sixth step obtains the seasonal component from these averages. The seasonal component is computed for each of the 61 sampling days in a year by finding the midmean of the yearly values for that sampling day. It is subtracted from block 4, leaving the Valley-wide smooth, and is placed in block 5. To obtain the values in Figure 1, we retransformed this seasonal component.

The smoothing algorithm 4253H is the following sequence of computations.[8] First, running medians of length 4 and 2 are applied to give

$$y_t^{(1)} = (1/2)\ \text{median}\ [y_{t-2}, y_{t-1}, y_t, y_{t+1}] + (1/2)\ \text{median}\ [y_{t-1}, y_t, y_{t+1}, y_{t+2}]. \quad (2)$$

Second, a running median of length 5 is applied to give

$$y_t^{(2)} = \text{median}\ [y_{t-2}^{(1)}, y_{t-1}^{(1)}, y_t^{(1)}, y_{t+1}^{(1)}, y_{t+2}^{(1)}]. \quad (3)$$

Third, a running median of length 3 is applied to give

$$y_t^{(3)} = \text{median}\,[y_{t-1}^{(2)}, y_t^{(2)}, y_{t+1}^{(2)}]. \tag{4}$$

Fourth, a running weighted average called hanning is applied to give

$$y_t^{(4)} = [y_{t-1}^{(3)} + 2\,y_t^{(3)} + y_{t+1}^{(3)}]/4. \tag{5}$$

The above formulas show that $y_t^{(4)}$ is obtained from 13 original data points, $y_{t-6}, \ldots, y_{t+6}$. To make the output the same length as the input, six points are joined to each end of the series. The points at the beginning are obtained by applying the sequence 4253H to the first 14 data points to give $y_7^{(4)}$ and $y_8^{(4)}$. The six new points, which are denoted by $y_{-5}, y_{-4}, \ldots, y_0$, are obtained by linear extrapolation:

$$y_j = y_7^{(4)} + (7-j)(y_7^{(4)} - y_8^{(4)}). \tag{6}$$

The points for the other end are obtained similarly.

SECTION 4

INTERPRETATION OF THE DATA

The decomposition of TSP and SS data allows comparison of the various components to possible causal factors. Some causal factors are regional and some are local; some vary rapidly and some vary slowly. Thus, the decomposition is useful because the causal factors relate to some components, but not to others.

SEASONAL COMPONENT

The seasonal component is not only a prominent feature of most environmental data, but often the component that is most difficult to explain unambiguously. This difficulty is due to the seasonal nature of most possible causes.

The TSP spring peak is interesting in that it seems to be related to annually recurring events in late March and early April. The most plausible explanation for this peak is regional agricultural and biological activity. This period is the planting season in the Tennessee Valley and also the season for release of pine pollen. Both of these particulate sources should be important at rural sites and quite possibly at industrial-urban monitoring sites as well.

The TSP and SS summer peaks coincide with many interrelated factors. These peaks result from the increased frequency of meteorological conditions conducive to the transformation, transport, and buildup of both primary and secondary pollutants. Among these factors are

- High incidence of stagnating anticyclonic (high-pressure) airmasses;
- High absolute atmospheric water vapor content;
- High insolation;
- High temperature;
- Higher convective and less advective mixing;
- Low frequency of regional rainfall; and
- Low wind speed.

Although anthropogenic emissions may be the origin of a significant portion of summer air pollution, the variation in emissions alone does not seem to account for these peaks since the power demand on the TVA system is as high in winter as in summer.

VALLEY-WIDE AND LOCAL SMOOTHS

Smooths are, by definition, representative of persistent behavior and, as such, are useful in determining the trend of the data. The Valley-wide smooth is indicative of persistent behavior common to all sites. The many features of the smooths shown in Figures 2 and 3 have not been analyzed, but two examples taken from the Valley-wide smooths and two examples from the local smooths will be discussed.

Examining the Valley-wide smooth for TSP in Figure 2, note that the fall of 1974 is a period with high levels. We attribute these unusually high levels to a prolonged dry spell. This dry spell shows the effect of meteorology on pollutant levels. The mechanisms for the pollutant increase are the dry conditions and the presence of stagnating high-pressure systems, which allow a greater amount of wind-borne soil and pollutant buildup.

Turning to the Valley-wide SS smooth in Figure 3, consider the general downward trend of the data. It appears that 1974 and 1975 experienced higher sulfate levels than did 1976 and 1977. What does this indicate? It could represent an actual decline in regional sulfate concentrations, which as mentioned previously, could be a function of year-to-year meteorological fluctuations. It also could be the result of the change in sampling techniques in July of 1976--the switch from Mine Safety Appliance Co. to Gelman Spectrograde high-volume filters. Subsequent experimentation with sulfate extraction suggests that the SS data obtained from the Gelman filters are on the average too low.

The local smooths for TSP and SS at Giles County in the fall of 1974 are unusually low. Examination of the TSP and SS data during this period indicates either extremely low pollutant concentrations or lost records. An examination of corresponding data collected from the nearby Cumberland Steam Plant indicated no unusual data. This suggests that this negative peak is due to a defective high-volume sampler.

The local SS smooth for LBL in August of 1976 is another interesting example. In this instance, the sixth-day sampling resulted in a smooth not typical of the entire month. Three of five sampling days during the month had high levels of SS. These levels, which were peculiar to LBL, were caused by transport from the Ohio Valley, a meteorological circumstance that occurs infrequently.[13] This particular peak, therefore, is not representative of the entire period.

This problem is an example of failure of the smoothing to separate the slowly varying and rapidly varying components. With sixth-day sampling, we are unable to separate these components. This problem, which is called aliasing[18] and is a form of confounding, makes explanation of the smooths more difficult. The effect of aliasing is further demonstrated in the following example.

The data used in this example are 192 days of daily SS values. In Figure 5, the six lines superimposed over the actual data are sixth-day smooths generated by using different starting days. Each smooth shows a

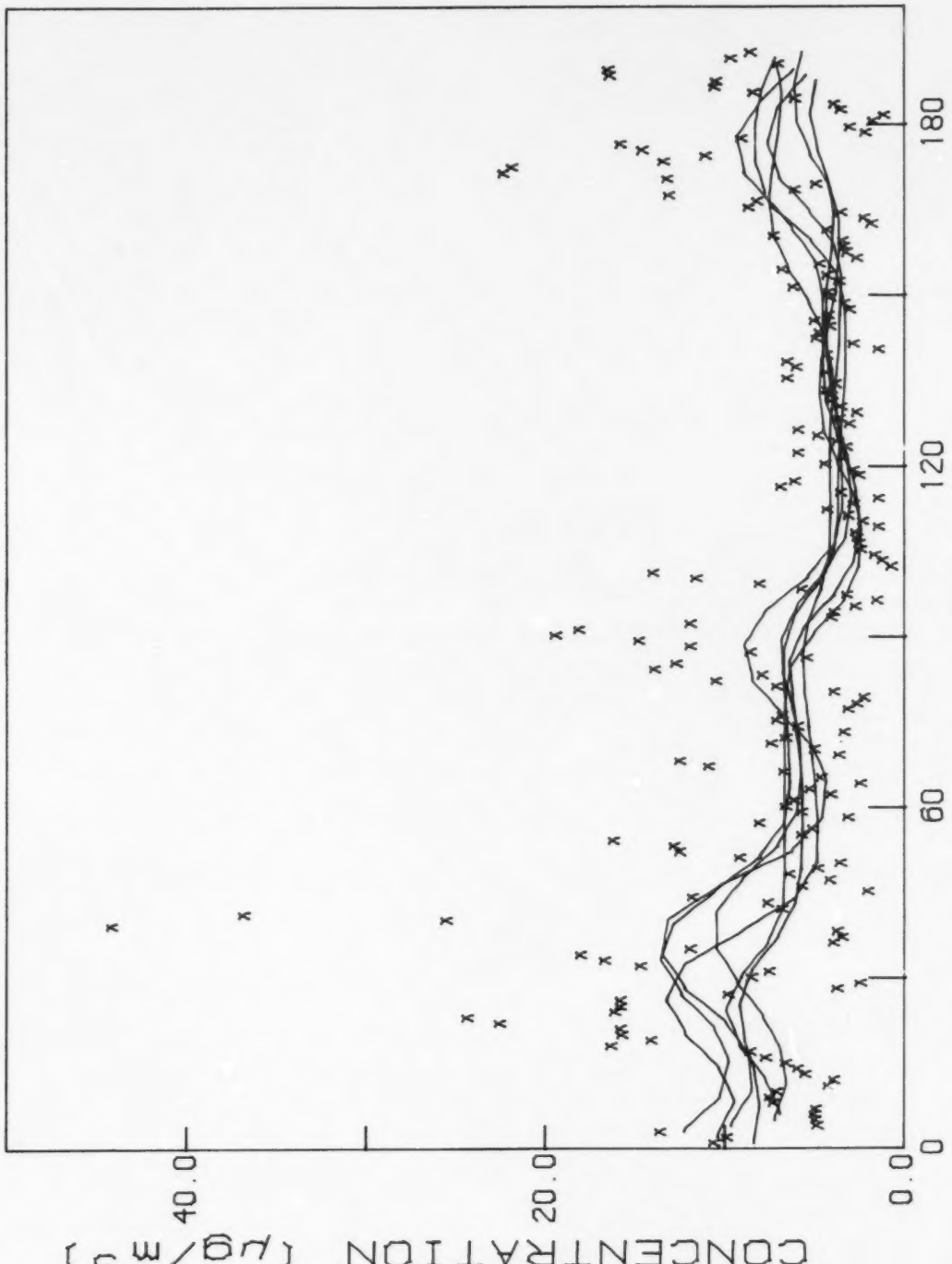

Figure 5. Daily sulfate data with sixth-day sampling smoothed.

single peak in late August or early September, although the actual data have three major peaks in this period. Thus, the six smooths do not describe the actual data very well. Further, they are not similar to each other. The problems of aliasing could be eliminated through daily sampling or at least reduced through more frequent sampling.

VALLEY-WIDE AND INDIVIDUAL ROUGHS

The roughs contain that part of each day's value unsupported by adjacent values. In other words, a rough is the irregular part of a series, the high-frequency component. The roughs serve as a means for detecting unusual episodes. Also, they can be used for comparisons among the high-frequency variations at the various sites.

Examination of the Valley-wide rough shows episodes that exhibit extreme levels throughout the region. Seventeen such episodes are considered in detail. Five episodes were chosen because they had the highest values of the SS Valley-wide rough. Of these, three are typical and two are unusual. The other twelve episodes have the four lowest values of the SS Valley-wide rough and the four highest and four lowest values of the TSP Valley-wide rough.

The episodes occurring on January 4, 1974, August 26, 1974, and July 4, 1975, are typical high-SS episodes. The common factor in these cases is the presence of a stagnating anticyclonic airmass. The winter episode of January 4, 1974, coincides with the presence of a cold polar continental anticyclone (PcK), which had been stagnating over middle-America since January 1. The summer episodes coincide with the presence of warm maritime anticyclones (TmW), which had stagnated over the southeastern United States. The presence of fog, smoke, haze, and low visibility are typically associated with such episodes.

The episodes occurring on January 28, 1974, May 22, 1974, May 7, 1976, and April 7, 1977, are typical low-SS episodes. The common factor associated with these cases is the presence of regional precipitation in substantial amounts on the days preceding and/or during the sampling day. This precipitation is associated with airmass convergence.

Episodes that do not fit into a similar mode require further investigation. The episodes on February 12, 1977, and on September 10, 1977, represent two such nontypical cases.

The episode occurring on February 12, 1977 is quite unusual. SS concentrations at the trend stations were 9.9, 6.5, 8.4, 8.4, and 7.6 $\mu g/m^3$, consistently above the winter seasonal mean of 4.5 $\mu g/m^3$. The meteorology on the sampling day was dominated by a cyclonic center, moving in a northeasterly direction across southeastern Missouri. The associated cold front resulted in measurable precipitation across the Valley on the day of sampling. The elevated SS values conceivably result from two factors: (1) static sampling during the five days before sampling, when the presence of a stagnating anticyclone could

have resulted in SS buildup, and (2) partial sampling of this stagnating airmass until the rain began. It is logical to assume that, had the cyclone not developed, SS concentrations would have been much greater.

The episode occurring on September 10, 1977, is also quite unusual. SS concentrations on this day were 16.9, 15.5, 30.4, 27.0, and 0.5 $\mu g/m^3$, generally above the summer seasonal mean of 10.4 $\mu g/m^3$. Indeed, as can be seen in the variation of values, this situation is unusual. On the day of sampling, the Valley meteorology was dominated by an approaching cold front from the northwest, followed by a maritime polar anticyclone (PmK). The 0.5-$\mu g/m^3$ concentration recorded at LBL is so unusually low that SS concentrations at nearby TVA steam plants were also checked. This check confirmed low SS concentrations in the northwest section of the Tennessee Valley--undoubtedly associated with the PmK airmass. The airmass to the southeast of the front is associated with much higher SS concentrations. Because of the complex meteorology on the days before sampling, resulting from the passage of a tropical depression, the origin of this prefrontal airmass is uncertain. The three-dimensional trajectory model of the National Weather Service Techniques Development Laboratory shows that on September 7, 8, and 9, the trajectories into the Valley were from the north to northeast. These trajectories crossed the large sulfur dioxide emissions sources in the Ohio Valley.

The episodes occurring on October 25, 1974, and July 28, 1975, are examples of one type of high-TSP episode. In this case, the common factor is a stagnating anticyclone, a PcK in the former episode and a TmW in the latter. The stagnating conditions are associated with fog, haze, smoke, low wind speed, and reduced visibility. We believe that, in cases such as these, fine particulates from natural and anthropogenic sources build up in the atmosphere and result in the elevated TSP levels.

The episodes occurring on January 4, 1974, and April 4, 1974, are also examples of high-TSP episodes. The mechanisms are, however, much different from the ones discussed above. In these cases, the episodes are associated with frontal activity, rain, and high wind speeds. We believe that in these cases, coarse particulates, primarily from natural sources, are carried aloft by the high winds associated with the frontal activity and result in the elevated TSP levels.

The episodes occurring on February 4, 1975, May 25, 1977, September 16, 1977, and November 3, 1977, are examples of low-TSP episodes. The common factor associated with these episodes is the presence of regional precipitation before and on the day of sampling. In all these episodes, the precipitation is associated with frontal activity. There is no readily apparent explanation for differentiating between meteorological conditions occurring during the second type of elevated TSP episodes and these low episodes. The differences are most likely related to the sources.

The individual roughs isolate unusual data points and describe for each site the rapidly varying part of sample-to-sample variation. Unusual data points may reflect an unusual set of environmental circumstances or an error in sampling, laboratory work, or recording. As such, the individual roughs may be used in quality assurance.

As seen in Table 1, the individual roughs are well correlated. This, indeed, is a manifestation of the common regional behavior. When compared with the Valley-wide rough, these correlations provide a quantitative measure of regional "representativeness." The Giles County site appears to be most representative of regional TSP behavior, whereas the Loudon, Hytop, and Giles County sites appear to be most representative of regional SS behavior.

SECTION 5

DESIGN OF MONITORING

Ambient monitoring can be used to estimate exposure as part of a study of pollution effects or to evaluate sources as part of a study of control strategies. To achieve this latter objective, power plant sources, other industrial sources, agricultural sources, other local sources, and distant sources must be resolved. Considerations important to accomplishing this are shown by the data analyzed in this paper.

One consideration is how to resolve the agricultural and biological contribution to the TSP. Compared with some other contributions, this contribution is believed to contain mostly larger-size particles and to be less dangerous to human health.[19] Whatever the relative health effects of various types of particles, this contribution must be distinguished as effectively as possible in studies of control strategies. Thus, studies of control strategies are an important basis for the frequently repeated recommendation that particulates be measured by size and chemical composition.

Another consid[illegible]on is how the meteorological influence can be removed. This inf[illegible] is important in the study of long-term variations, which is one [illegible] TVA's purposes for monitoring at these isolated sites. How such variations apparent in the data are interpreted depends on their cause: Variations caused by the weather have different implications for control than variations caused by other factors. Thus, long-term variations must be analyzed by removing the influence of year-to-year differences in the weather to obtain the series that would have occurred had each year's weather been the same. This series should show the part of the variation caused by changes in emissions.

The meteorological influence is also likely to be important in analyzing data from sites surrounding a power plant. This analysis could start with the same decomposition used above. Because all the sites would be sampled on the same day, the common rough, which is the analog of the Valley-wide rough, would be subtracted from the individual roughs to obtain a local rough for each site. The dependence of the local roughs and smooths on plume behavior would contain the evidence of pollution from the power plant. However, this dependence might exist even with no power plant contribution because of other sources. Thus, the resolution of sources also arises in this context, showing that analysis involving the weather will also be important for power plant data.

Analysis involving the meteorological influence, although it is never easy, is made harder by the aliasing problem. In the analysis of these data, the aliasing problem prevents separation of slowly varying components from rapidly varying components. Such separation is important in an observational study such as this, where the objective is to explain as much of the variation as possible. If the sampling were daily, the data would be separated into more than just rough and smooth components. The most irregular component would contain rare meteorological events as

well as the results of measurement blunders. Another component would reflect mostly the passage of weather systems, thus tracking the day-to-day variations in transport. A third component would be compared with monthly summaries of causal factors in the same way that we would like to compare Figures 2 and 3 with such summaries. This more extensive decomposition should allow monitoring to provide, under some circumstances, better information than modeling.

The recommendation that the sampling frequency for particulates be increased has been made previously on the basis that particulate measurements are a random sample.[20] Although this basis for thinking about air quality data is widespread,[21] it fails to acknowledge the possibility of modeling and adjusting for meteorological and other influences. When adjustment for these influences is considered, the major problem with sixth-day sampling is seen to be aliasing rather than accuracy.

The features of these data, revealed by the analysis demonstrated in this paper, suggest various changes to be made in the data collection. These changes include more resolution in the sampling itself and collection of more ancillary data. If these changes were made, adequate data for more detailed and quantitative model building would become available. The analysis demonstrated here has thus been shown to be important in ensuring that all the data necessary to satisfy the purposes of the monitoring are collected. Analysis with this purpose should be a part of all ongoing monitoring.

REFERENCES

1. Jutze, G. A., and K. E. Foster. Recommended Standard Method for Atmospheric Sampling of Fine Particulate Matter by Filter Media--High Volume Sampler. J. Air Pollut. Control Assoc., 17:17-25, 1967.

2. U.S. Public Health Service. Determination of Sulfate in Atmospheric Suspended Particulates. 999-AP-11, 1965.

3. Appendix B--Reference Method for the Determination of Suspended Particulates in the Atmosphere (High Volume Method). Fed. Regist., 36(84):8191-8194, 1971.

4. U.S. Environmental Protection Agency. Tentative Method for the Determination of Sulfates in the Atmosphere (Automated Technicon II Methylthymol Blue Procedure), 1977.

5. Goldsmith, B. J., and J. R. Mahoney. Implications of the 1977 Clean Air Act Amendments for Stationary Sources. Environ. Sci. Technol., 12:144-149, 1978.

6. Pratt, J. W., et al. Environmental Monitoring. National Academy of Sciences, Washington, D.C., 1977.

7. Tukey, J. W. Exploratory Data Analysis. Addison-Wesley, Reading, Mass., 1977.

8. Velleman, P. F. Robust Nonlinear Data Smoothers: Definitions and Recommendations. Proc. Natl. Acad. Sci. USA, 74:434-436, 1977.

9. Hidy, G. M., E. Y. Tong, and P. K. Mueller. Design of the Sulfate Regional Experiment (SURE), vol. 1. EPRI EC-125, Electric Power Research Institute, 1976.

10. Lioy, P. J., G. T. Wolff, J. S. Czachor, P. E. Coffey, W. N. Stasiuk, and D. Romano. Evidence of High Atmospheric Concentrations of Sulfates Detected at Rural Sites in the Northeast. J. Environ. Sci. Health, A12:1-14, 1977.

11. Galvin, P. J., P. J. Samson, P. E. Coffey, and D. Romano. Transport of Sulfate to New York State. Environ. Sci. Technol., 12:580-584, 1978.

12. Tong, E. Y., and R. B. Batchelder. Compilation and Analysis of Data Sets for the Evaluation of Regional Sulfate Models. Teknekron, Inc., Berkeley, California, 1978.

13. Reisinger, L. M., and T. L. Crawford. August 1976 Sulfate Episodes in the Tennessee Valley Region. TVA/EP-79/04, Tennessee Valley Authority, Chattanooga, Tennessee.

14. U.S. Environmental Protection Agency. National Air Quality and Emissions Trend Report, 1976. EPA-450/1-77-002, 1977.

15. Altshuller, A. P. Atmospheric Sulfur Dioxide and Sulfate--Distribution of Concentration in Urban and Nonurban Sites in the United States. Environ. Sci. Technol., 7:709-712, 1973.

16. Rowe, M. D., S. C. Morris, and L. O. Hamilton. Potential Ambient Standards for Atmospheric Sulfates: An Account of a Workshop. J. Air Pollut. Control Assoc., 28:772-775, 1978.

17. Gnanadesikan, R. Methods for Statistical Data Analysis of Multivariate Observations. John Wiley and Sons, Inc., New York, 1977. p. 132.

18. Bloomfield, P. Fourier Analysis of Time Series: An Introduction. John Wiley and Sons, Inc., New York, 1976.

19. Hidy, G. M., et al. Summary of the California Aerosol Characterization Experiment. J. Air Pollut. Control Assoc., 25:1106-1114, 1975.

20. Tong, E. Y., and S. A. DePietro. Sampling Frequencies for Determining Long-Term Average Concentrations of Atmospheric Particulate Sulfates. J. Air Pollut. Control Assoc., 27:1008-1011, 1977.

21. Mage, D. T., and W. R. Ott. Refinements of the Lognormal Probability Model for Analysis of Aerometric Data. J. Air Pollut. Control Assoc., 28:796-798, 1978.